NICEATM

National Toxicology Program Interagency Center for the Evaluation Of Alternative Toxicological Methods

ICCVAM

Interagency Coordinating Committee on the Validation of Alternative Methods

Preliminary Evaluation of the Underprediction Rate of the *In Vivo* Dermal Irritation Test Method Part I: Introduction

William S. Stokes, D.V.M., D.A.C.L.A.M.
Director, NICEATM

Scientific Advisory Committee on Alternative Toxicological Methods
October 20, 2004
Research Triangle Park, NC

National Toxicology Program

NIEHS

NATIONAL INSTITUTES OF HEALTH

DEPARTMENT OF HEALTH & HUMAN SERVICES • USA •

Acknowledgements

ICCVAM Dermal Corrosivity and Irritation Working Group (DCIWG)

Marilyn Wind, CPSC
Kailash Gupta, CPSC
Barnett Rattner, DOI
Steve Hwang, DOT
George Cushmac, DOT
Angela Auletta, EPA
Masih Hashim, EPA
Leonard Keifer, EPA
Marianne Lewis, EPA
John Redden, EPA
Mark Perry, EPA
Amy Rispin, EPA (Co-Chair)
Hari Mukhoty, EPA
Karen Hamernik, EPA
Debbie McCall, EPA

Leonard Schechtman, FDA
Robert Bronaugh, FDA
Jill Merrill, FDA
Donnie Lowther, FDA
Abby Jacobs, FDA (Co-Chair)
Vera Hudson, NLM
William Stokes, NIEHS
William Eastin, NIEHS
Buck Grissom, NIEHS
Rajendra Chhabra, NIEHS
Surender Ahir, OSHA

Acknowledgements

- Dr. Joe Haseman, NIEHS/NTP Biostatistician (Retired)

- NICEATM Staff
  - Dr. David Allen, ILS Inc.
  - Dr. Neepa Choksi, ILS Inc.
  - Dr. Ray Tice, ILS Inc.

- Data Source
  - European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) *Skin Irritation and Corrosion: Reference Chemicals Data Bank.* Technical Report No. 66. March, 1995. Brussels, Belgium

Outline

- Introduction
  - Background
  - Current Testing Procedures
  - Prior Analyses
  - Study Objectives
  - Database
  - Future Plans
- Data Analysis

Background

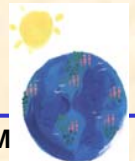
- **Draize rabbit skin test method**
  - **Used since the 1940's to identify skin irritants and corrosives**

- **Skin corrosion**: ***the production of irreversible damage to skin following application of a test substance for up to 4 hrs***

- **Skin irritation**: ***the production of reversible damage to skin following application of a test substance for up to 4 hours***

Background

- 2003 Globally Harmonized System of Classification and Labelling of Chemicals(GHS)
  - Tiered testing approach incorporating the use of valid and accepted *in vitro* methods for dermal irritation should be considered
- Non-animal alternative methods proposed for assessing dermal irritation
  - EPISKIN™, EpiDerm™, and SIFT
  - ECVAM validation in progress
    - ICCVAM and NICEATM liaisons
- Estimates of the underprediction likely in an animal would assist with interpreting the usefulness and limitations of *in vitro* test methods



Tiered-Testing Strategy

Valid and accepted *in vitro* dermal corrosion test → Positive Response → Classify as corrosive

Negative Response or no data

Valid and accepted *in vitro* dermal irritation test* → Positive Response → Classify as irritant

Negative response or no data

***In vivo* dermal corrosion (1 animal)** → Corrosive response → Classify as corrosive

Negative response

***In vivo* dermal irritation test (3 animals total)** → Irritant response → Classify as irritant

Negative response → No further testing → Classify as non-irritant

When is it ethical to perform human patch testing? → Irritant response → Classify as irritant

Not as above → Non-irritant response → No further testing

***Must be capable of detecting false negative chemicals from an *in vitro* corrosivity test.**

ICCVAM
NICEATM

Current Testing Procedures

- Draize rabbit skin test method
- Current test guideline procedures since 1981 (OECD TG 404)
- Test method protocol
  - 0.5 mL or 0.5 g of test substance applied to intact skin with patch for 4 hours
    - originally 6 animals; reduced to 1-3 animals in 1992
    - Test substance removed after 4 hr exposure period
  - Erythema and edema scored at 24, 48, and 72 hours
  - Observation for 14 days to determine persistence or delayed effects

Dermal Irritation Scoring

- **Erythema**
  - 1 = Very slight (barely perceptible)
  - 2 = Well defined
  - 3 = Moderate to severe
  - 4 = Severe erythema (beefy redness) to eschar formation preventing grading of erythema

- **Edema Scores**
  - 1 = Very slight (barely perceptible)
  - 2 = Slight (edges of area well defined by definite raising)
  - 3 = Moderate (raised approximately 1 mm)
  - 4 = Severe (raised more than 1 mm and extending beyond area of exposure)

Hazard Classification for Dermal Irritation

- UN Globally Harmonized System (GHS), 2003
- Classification Scheme
  - Irritant
    - At least 2 animals have an average erythema *or* edema score that is greater than 2.3
  - Mild irritant
    - At least 2 animals have an average erythema *or* edema score that is between 1.5 and 2.3
  - Nonirritant
    - If no more than 1 animal has an average erythema or edema score that is greater than 1.5

Prior Analysis of the Reproducibility of the Rabbit Dermal Irritation Test

- Weil and Scala (1971)
  - Evaluated the reproducibility of the Draize rabbit skin test method within and among 24 laboratories for 10 substances
- This study is the only formal evaluation of the reproducibility of the Draize rabbit skin test method
- Conclusions
  - Moderate intra-laboratory reproducibility
  - Low inter-laboratory reproducibility
  - Primary reasons for the low inter-laboratory reproducibility attributed to the subjective nature of the visual observations and variations in procedures among labs

Weil CS, Scala RA. 1971. Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. Toxicol. App. Pharmacol. 19:276-360.

Limitations of the Weil and Scala Analysis

- The standard protocol used was different from the current Draize *in vivo* rabbit skin test method protocol in use since 1981
  - The Weil and Scala studies used a 24-hour exposure period versus the current maximum 4-hour exposure
  - Prolonged exposure likely responsible for corrosive lesions observed for several irritants
- Good Laboratory Practice (GLP) Guidelines had not yet been established
  - Impact unknown

Study Objectives

- **Evaluate ECETOC Chemical Data Bank to estimate the likelihood of underpredicting:**
  - **An irritant as a mild irritant**
  - **An irritant as a non-irritant**
  - **A mild irritant as a non-irritant**

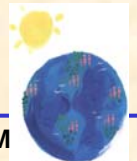
- **Data may assist in decisions on acceptable false-negative rate for irritant effects for *in vitro* test methods proposed as complete replacements for the rabbit skin test**
  - **i.e., those tests where no *animal* testing would be performed and *in vitro* results would serve as the basis for hazard classification and labeling**

In Vivo Dermal Irritation Database

- ECETOC Reference Chemicals Data Bank
  - 164 chemicals in 197 studies
  - Represent a wide range of chemical classes
  - Studies were performed according to OECD TG 404 and GLPs
  - 23 chemicals were tested in multiple studies
  - Most chemicals tested in 3-6 animals

Source	Number of Animals Used per Study					
	1	2	3	4	5	6
ECETOC[1]	1	0	96	90	0	10

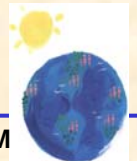
[1]European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC), Skin Irritation and Corrosion: Reference Chemicals Data Bank. Technical Report No. 66. Belgium. (All studies followed OECD TG 404 and GLP Guidelines)



ICCVAM
NICEATM

Future Analysis Plans

- **Continue to seek high quality test data to add to the database:**
  - ***Federal Register* Notice (July 16, 2003)**
    - **Requested *in vivo* dermal data for chemicals that could be considered for reference chemicals**
  - **EPA TSCATS database**
    - **Current collaboration with EPA OPPTS to obtain reports for ~2400 commercially available chemicals with dermal test results**
    - **638 reports reviewed to date, *but*:**
      - **Limited individual animal data provided**
      - **Many studies were conducted prior to 1981 (exposure of 24 hr vs. 4 hr)**
- **Perform reanalysis when EPA data review completed**



NICEATM

National Toxicology Program Interagency Center for the Evaluation Of Alternative Toxicological Methods

ICCVAM

Interagency Coordinating Committee on the Validation of Alternative Methods

Preliminary Evaluation of the Underprediction Rate of *In Vivo* Dermal Irritation Test Method Part II: Data Analysis

Joseph Haseman, Ph.D.

Scientific Advisory Committee on Alternative Toxicological Methods
October 20, 2004
Research Triangle Park, NC

National Toxicology Program

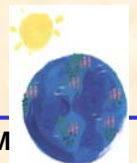
NIEHS

NATIONAL INSTITUTES OF HEALTH

DEPARTMENT OF HEALTH & HUMAN SERVICES • USA •

Definition of Underprediction Rate

- The under-prediction rate of an irritation test is defined as the probability that an irritant substance will not be classified as an irritant when subjected to the test
  - e.g., it will produce responses that classify an irritant as a non-irritant in the rabbit model
- The underprediction rate depends on
  - the distribution of animal responses for substances assigned to a specific classification category
  - the strategy that is used to assign a test substance to a classification category



Classification of Potential Outcomes

Erythema or Edema Score			Classification	Probability Calculation
<1.5	1.5-2.3	>2.3		
3	0	0	Negative	$(P_N)^3$
2	1	0	Negative	$3P_N^2P_M$
2	0	1	Negative	$3P_N^2P_I$
1	1	1	Mild Irritant	$6P_NP_MP_I$
1	2	0	Mild Irritant	$3P_M^2P_N$
0	3	0	Mild Irritant	$(P_M)^3$
0	2	1	Mild Irritant	$3P_M^2P_I$
1	0	2	Irritant	$3P_I^2P_N$
0	1	2	Irritant	$3P_I^2P_M$
0	0	3	Irritant	$(P_I)^3$

P_N: probability that erythema/edema score<1.5; P_M: score = 1.5-2.3, P_I: score > 2.3

ICCVAM

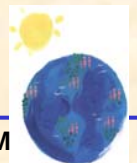
NICEATM

Calculation of the Underprediction Rate

- The distribution of animal responses for each irritancy class (i.e., irritant, mild irritant, nonirritant) was calculated
- Using this distribution and the possible outcomes provided in the previous table, response probabilities were calculated for each outcome for a specific irritancy classification.
- For each irritancy classification, these probabilities were then summed to provide an overall classification likelihood.
- 2 approaches were used:
  1) All substances in the database were used, OR
  2) Only substances tested multiple times were used

Distribution of Animal Scores (Approach 1)

Estimated Probability of … (No. animals)	True Classification of Test Substance		
	Nonirritant	Mild Irritant	Irritant
An animal scoring < 1.5	95.7% (222)	14.2% (47)	0.7% (1)
An animal scoring 1.5 - 2.3	3.9% (9)	81.6% (270)	19.2% (28)
An animal scoring > 2.3	0.4% (1)	4.2% (14)	80.1% (117)
No. Studies Evaluated	*66*	*88*	*43*



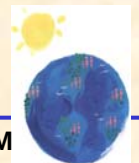
Example Calculation of Probability - Likelihood of a Nonirritant being Classified as a Nonirritant

Erythema or Edema Score			Classification	Probability Calculation
< 1.5	1.5 - 2.3	> 2.3		
3	0	0	Negative	$(P_N)^3$
2	1	0	Negative	$3P_N^2P_M$
2	0	1	Negative	$3P_N^2P_I$

$(P_N)^3 + 3P_N^2P_M + 3P_N^2P_I = (0.957)^3 + [3(0.957)^2(0.039)] + [3(0.957)^2(0.004)] = 0.995 = 99.5\%$

Estimated Probabilities of Classification (Approach 1)

		True Classification of Test Substance		
		Negative	Mild Irritant	Irritant
Our Classification of Test Substance	Negative	99.5%	5.5%	0.01%
	Mild Irritant	0.5%	94.0%	10.3%
	Irritant	<0.01%	0.5%	89.7%



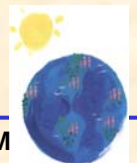
Distribution of Animal Scores (Approach 2)

Estimated Probability of … (No. animals)	True Classification of Test Substance		
	Nonirritant	Mild Irritant	Irritant
An animal scoring < 1.5	91.7% (55)	11.6% (13)	0% (0)
An animal scoring 1.5 - 2.3	8.3% (5)	79.5% (89)	42.4% (14)
An animal scoring > 2.3	0% (0)	8.9% (10)	57.6% (19)
No. Chemicals Evaluated	*8*	*12*	*3*

Estimated Probabilities of Classification (Approach 2)

		True Classification of Test Substance		
		Negative	Mild Irritant	Irritant
Our Classification of Test Substance	Negative	98.0%	3.7%	0%
	Mild Irritant	2.0%	94.0%	38.7%*
	Irritant	0%	2.2%	61.3%

*Database includes only 3 irritants



Estimated Underprediction Rates of the *In Vivo* Dermal Irritation Test Method

Outcome	Approach 1*	Approach 2*
Underprediction of Irritant as Mild Irritant	10.3%	38.7%**
Underprediction of Irritant as Negative	0.01%	0%
Underprediction of Mild Irritant as Negative	5.5%	3.7%
Underprediction of Irritant and Mild Irritant as Negative	5.5%	3.7%

*Approach 1 = All chemicals used; Approach 2 = Only multiply-tested chemicals

**Database includes only 3 irritants

ICCVAM

NICEATM

Mean Scores for the 3 Multiply Tested Skin Irritants

Chemical (Study No.)	Mean Erythema					Mean Edema				
	An. 1	An. 2	An. 3	An. 4	Study Mean	An. 1	An. 2	An. 3	An. 4	Study Mean
Alpha-terpineol (1)	1.7	2.0	2.3	-	2.0	2.0	2.3	3.0	-	2.4
Alpha-terpineol (2)	2.0	2.7	2.0	2.0	2.2	3.0	3.0	2.7	1.7	2.6
Alpha-terpineol (3)	2.0	2.0	1.7	2.0	1.9	2.7	2.0	0.7	3.0	2.1
Cyclamen aldehyde (1)	2.7	2.0	2.0	-	2.2	3.0	3.0	2.7	-	2.9
Cyclamen aldehyde (2)	2.0	2.0	2.0	2.0	2.0	2.3	2.7	2.0	1.7	2.2
Cyclamen aldehyde (3)	2.0	2.0	2.0	2.7	2.2	2.7	2.7	2.3	3.0	2.7
Cyclamen aldehyde (4)	2.0	2.0	2.0	2.0	2.0	1.3	1.0	2.0	1.3	1.4
Lilestralis/Lilial (1)	1.7	2.0	2.3	-	2.0	2.0	2.7	3.0	-	2.6
Lilestralis/Lilial (2)	2.0	1.7	2.0	2.0	1.9	1.7	1.7	2.3	1.0	1.7

Conclusions

- Within the limits of the assumptions, the under-prediction of:
  - an irritant as a mild irritant ranged from 10.3% to 38.7%*
  - an irritant as a nonirritant ranged from 0% to 0.01%
  - a mild irritant as a nonirritant ranged from 3.7% to 5.5%
- Based on these data, the likelihood that an irritant would be misclassified as a nonirritant is less than 0.01%.
- The relatively small number of irritants among the multiply-tested substances may impact the reliability of the estimated underprediction rate.

*The 38.7% underprediction rate is based on only 3 irritants